

FIR AND IIR TECHNIQUES FOR BANDLIMITED EXTRAPOLATION OF SEQUENCES

P. P. VAIDYANATHAN and VINAY P. SATHÉ

Dept. of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

ABSTRACT

The problem of extrapolating a finite length sequence into a bandlimited sequence has been formulated in the past in terms of pseudoinverse techniques. Iterative and non iterative algorithms for the extrapolation have also been proposed. All these algorithms, however, attempt to produce an ideally bandlimited answer, which, in reality, is infinitely long and noncausal, and cannot be generated in finite time. In this paper the problem is converted to one of finding a finite length input to a practical FIR or IIR filter in such a way that a segment of the output matches the signal to be extrapolated. The fact that the generated signal is the output of a filter ensures that it is bandlimited to an extent determined largely by the filter quality. The advantage of this viewpoint is that we can exploit the rich literature which is available for efficient filter implementation (such as multistage and IFIR techniques, and efficient IIR techniques based on allpass building blocks).

I. INTRODUCTION

Given a finite length sequence $x(n)$, $0 \leq n \leq N-1$, suppose we wish to find a σ -bandlimited (abbreviated σ -BL) sequence $y(n)$ such that

$$y(n+M) = x(n), \quad 0 \leq n \leq N-1, \quad (1)$$

where M is an arbitrary finite integer. This is called the discrete-time σ -BL extrapolation problem [1-3]. The term σ -BL implies that $Y(e^{j\omega}) = 0$ for $|\omega| \geq \sigma$. The historical predecessor for this is the continuous-time version, which is discussed in [4] and references therein.

1.1. On existence and nonuniqueness.

In the continuous time case the problem does not necessarily have a solution i.e., there may or may not exist a bandlimited extrapolation; and if does exist it is unique

Work supported in parts by the NSF grants DCI 85 52579 and MIP 86 04456, and by Caltech's Programs In Advanced Technology grant sponsored by Aerojet General, General Motors, and TRW.

[4]. For the discrete time case, in contrast, there always exists a solution and it is non-unique [1],[10]; in fact there is an infinite number of solutions. (In most known techniques the solution is rendered unique by introducing the constraint that $y(n)$ have smallest possible energy.) We shall now briefly justify these statements. Discrete time σ -BL sequences are not subject to many of the restrictions which a continuous time bandlimited signal has to satisfy. For example it is possible to have a σ -BL sequence $t(n)$ which is zero for an arbitrary number of consecutive sample numbers. To see this suppose $s(n)$ is some (nonzero) σ -BL sequence. If we pass $s(n)$ through an FIR filter $G(z) = \sum_{n=0}^{N-1} g(n)z^{-n}$, the samples of the output $t(n)$ for $0 \leq n \leq N-1$ are given by

$$\mathbf{t} = \mathbf{S}\mathbf{g} \quad (2)$$

where

$$\mathbf{t} = [t(0) \ t(1) \ \dots \ t(N-1)]^T$$

$$\mathbf{S} = \begin{bmatrix} s(0) & s(-1) & \dots & s(-N+1) \\ s(1) & s(0) & \dots & s(-N+2) \\ \vdots & \vdots & \ddots & \vdots \\ s(N-1) & s(N-2) & \dots & s(0) \end{bmatrix} \quad (3)$$

$$\mathbf{g} = [g(0) \ g(1) \ \dots \ g(N-1)]^T.$$

The $N \times N$ matrix above is easily ensured to be nonsingular, for example by taking $s(n)$ to be the famous sinc-sequence, i.e., $s(n) = \sin \sigma n / \pi n$. So we can solve for the filter coefficients $g(n)$ which ensure that the output samples $t(0) \dots t(N-2)$ are zero. At the same time $t(n)$ is σ -BL because $s(n)$ is σ -BL. Of course we can interchange the roles of $G(z)$ and $s(n)$ and say that $t(n)$ has been generated as the output of an ideal lowpass filter $s(n)$ by applying a finite length input $g(0), g(1) \dots g(N-1)$.

Suppose we have generated a set of σ -BL sequences $y_k(n)$, $0 \leq k \leq N-1$ satisfying

$$y_k(n) \begin{cases} = 0, & 0 \leq n \leq k-1 \\ \neq 0 & n = k. \end{cases} \quad (4)$$

In particular note that $y_0(0) \neq 0$. It is then clear that we can find a linear combination $y(n) = \sum_{k=0}^{N-1} \alpha_k y_k(n)$ such

that (1) holds with $M = 0$. We find α_k recursively from

$$\alpha_k = [x(k) - \sum_{i=0}^{k-1} \alpha_i y_i(k)] / y_k(k) \quad (5)$$

because of the triangular set of equations induced by (4). Evidently $y(n)$ is σ -BL because it is a sum of N σ -BL signals. This proves that we can, in principle, trivially find a σ -BL extrapolation. If we now add an arbitrary σ -BL signal $r(n)$ to the result $y(n)$ with the property that $r(n) = 0$ for $0 \leq n \leq N-1$, the result continues to be a valid σ -BL extrapolation! These arguments prove the existence and non uniqueness of σ -BL extrapolations.

1.2. The lowpass filtering viewpoint.

The above procedure, conceptually, had two steps. We first generate a set of N σ -BL waveforms as outputs of a lowpass filter and then we combine these N waveforms. These two steps can be lumped into one as follows: we shall generate the σ -BL extrapolation $y(n)$ as the output of an ideal lowpass filter in response to a finite length input $u(0), u(1), \dots, u(N-1)$. The N degrees of freedom in the choice of input are sufficient to force the N conditions (1) on the output $y(n)$ (for some M , say $M = 0$). Evidently $y(n)$ remains bandlimited regardless of the values of $u(n)$. Since $y(n)$ is the convolution of $u(n)$ with the ideal lowpass impulse response $\sin(\sigma n)/\pi n$, we can compute the input samples $u(n)$ from

$$\mathbf{x} = \hat{\mathbf{L}} \mathbf{u} \quad (6)$$

where $\hat{\mathbf{L}} = [L_{mn}]$ is the truncated lowpass matrix, (i.e., $L_{mn} = \sin \sigma(m-n)/\pi(m-n)$), and

$$\begin{aligned} \mathbf{x} &= [x(0) \quad \dots \quad x(N-1)]^T \\ \mathbf{u} &= [u(0) \quad \dots \quad u(N-1)]^T. \end{aligned} \quad (7)$$

The extrapolated output can then be computed simply by passing $u(n)$ through the ideal filter $s(n)$ (which is doubly infinitely long and unstable).

It turns out that the approach in [1] (based on pseudoinverses) gives rise to the same result as this. To see this connection, let us define the matrix

$$\mathbf{S} = [\mathbf{0} \quad \mathbf{I}_N \quad \mathbf{0}]. \quad (8)$$

This has N rows and infinite number of columns, with the identity matrix \mathbf{I}_N occupying columns 0 through $N-1$. This is merely a truncation matrix, which operates on a infinite sequence and retains the N samples numbered 0 through $N-1$. Next define the ideal lowpass matrix \mathbf{L} as $\mathbf{L} = [L_{mn}]$, $-\infty \leq m, n \leq \infty$. With the matrices \mathbf{L}, \mathbf{S} and $\hat{\mathbf{L}}$ defined like this, we have the relation $\hat{\mathbf{L}} = \mathbf{S} \mathbf{L}^T$.

From the above discussions the finite length input $u(n)$ should then be chosen (for a given segment $x(n)$ of length N) as $\mathbf{u} = \hat{\mathbf{L}}^{-1} \mathbf{x}$ so that the output of the ideal lowpass filter (the extrapolated signal) is

$$\mathbf{y} = \mathbf{L} \mathbf{S}^T \mathbf{u} = \mathbf{L} \mathbf{S}^T (\mathbf{S} \mathbf{L}^T)^{-1} \mathbf{x}. \quad (9)$$

This output, by construction of $u(n)$, satisfies (1) with $M = 0$. This can be written (using the fact $\mathbf{L}^T = \mathbf{L}^2 = \mathbf{L}$) in the form

$$\mathbf{y} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{x} = \mathbf{A}^\# \mathbf{x}, \quad (10)$$

where $\mathbf{A} = \mathbf{S} \mathbf{L}^T$. But (10) is nothing but the minimum norm least squares (MNLS) solution [7] for the problem of approximating \mathbf{x} with $\mathbf{A} \mathbf{y} (= \mathbf{S} \mathbf{L} \mathbf{y})$, where \mathbf{x} is finite dimensional and \mathbf{y} is infinite dimensional. This proves that $y(n)$ has least energy among all σ -BL extrapolations.

II. APPROACHES BASED ON FIR AND IIR FILTERING

In the above approach, the extrapolated signal is doubly infinitely long, and is band limited in the ideal sense. In practice, however, one has to limit the length of $y(n)$ in some way. Truncation of $y(n)$ would normally lead to the Gibbs phenomenon so an obviously better method would be to multiply $y(n)$ with a window. The result, however is neither bandlimited, nor optimal in any specific sense.

2.1. Previous work

A more systematic approach would be to reformulate the problem taking into account the finite length of $y(n)$. For example, we can try to find $y(n)$ of length $L > N$ such that a segment of length N satisfies (1), with the remaining samples chosen to minimize the out of band energy

$$\phi_1 = \int_{\sigma}^{\pi} |Y(e^{j\omega})|^2 d\omega. \quad (11)$$

(For complex sequences, the integral should also include negative frequencies; these generalizations are trivial). The usual result of this is that $y(n)$ tends to have much higher energy than $x(n)$. This difficulty is overcome by minimizing an objective function

$$\phi = \alpha \phi_1 + (1 - \alpha) \phi_2, \quad 0 < \alpha < 1, \quad (12)$$

where $\phi_2 = \sum |y(n)|^2$. This has been done in [5,6]. While this approach gives rise to optimal finite length extrapolations in the sense of minimizing (12), it is computationally very expensive [5], as it involves inversion of a $K \times K$ matrix (with no special structure such as the Toeplitz property to be exploited) where $K = L - N$. In practice K can

be much larger than N . We shall now develop techniques for suboptimal extrapolations which are more efficient.

Notice, in any case, that the problem has now been reduced to an FIR filter design problem! We wish to find the impulse response $y(n)$ of a lowpass filter with minimum stopband energy. There are no 'passband requirements', but instead we have the time-domain constraint (1). This also opens up more possibilities: for example, we can try to use a Remez exchange type of algorithm (or linear programming) to minimize the maximum magnitude of $Y(e^{j\omega})$ in the region $\sigma \leq \omega \leq \pi$ under the constraint (1). This can be considered as a fanciful 'minimax, finite-length σ -BL extrapolation problem'.

2.2. The new approach

Let $h(n)$ be the impulse response of a causal lowpass filter with transfer function $H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}$ where $h(0) \neq 0$. Our aim is to find a causal input $u(n)$, $0 \leq n \leq M+N-1$, of finite length $M+N$ such that the samples $y(M), \dots, y(M+N-1)$ of the output $y(n)$ agree with the segment $x(k)$, $0 \leq k \leq N-1$ (see (1)). Note that we permit the input sequence to be longer than the segment length N , by an amount M . The purpose of M will soon be explained. The constraint (1) can be written as

$$\mathbf{x} = \mathbf{A}_M \mathbf{u}_M, \quad (13)$$

where \mathbf{x} is as in (7),

$$\mathbf{u}_M = [u(0) \ u(1) \ \dots \ u(M+N-1)]^T, \quad (14)$$

and

$$\mathbf{A}_M = \begin{bmatrix} h(M) & \dots & h(0) & 0 & \dots & 0 \\ h(M+1) & \dots & h(1) & h(0) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h(M+N-1) & \dots & \dots & \dots & \dots & h(0) \end{bmatrix}. \quad (15)$$

Eqn. (13) represents N equations in $N+M$ variables, and there exists \mathbf{u}_M which satisfies (13) exactly. To see this note that the matrix \mathbf{A}_M has full rank N because the rightmost $N \times N$ triangular submatrix is nonsingular (since $h(0) \neq 0$). For $M > 0$ there exists more than one solution. From the theory of Moore Penrose pseudoinverses [7] we know that for any M there exists a unique solution \mathbf{u}_M with minimum norm (MNLS solution) given by

$$\hat{\mathbf{u}}_M = \mathbf{A}_M^\# \mathbf{x}, \quad (16)$$

where $\mathbf{A}_M^\#$ is the Moore Penrose pseudoinverse of \mathbf{A}_M , which in the present case is given by $\mathbf{A}^\#$ in (10).

What happens if we replace M with $M+1$? The new set of equations is

$$\mathbf{x} = \mathbf{A}_{M+1} \mathbf{u}_{M+1}, \quad (17)$$

where \mathbf{A}_{M+1} is defined in an obvious manner. Evidently $\mathbf{u}_{M+1} = [0 \ \hat{\mathbf{u}}_M^T]^T$ is a solution to (17), but is not the MNLS solution. The MNLS solution, given by $\hat{\mathbf{u}}_{M+1} = \mathbf{A}_{M+1}^\# \mathbf{x}$, has smaller norm than $\hat{\mathbf{u}}_M$. As we solve (13) for successively larger M , the energy of the input therefore gets smaller and smaller. As the filter tends to attenuate the out-of-band energy of $u(n)$ anyway, the sequence $u(n)$ tends to look more and more like a lowpass signal. This intuitive expectation will be justified by examples. The output signal satisfies (1) exactly, and of course has very little energy for $|\omega| > \sigma$, due to the lowpass filtering action.

Reason for improved efficiency. Notice that once we solve for the samples of the input $u(n)$, we can compute all samples of the extrapolated output $y(n)$ by using any technique for digital filter implementation. The filter $h(n)$ can be FIR or IIR. In the FIR case, one can use some of the recent techniques [8],[9], for efficient computation of $y(n)$. For the IIR case, the IIR nature automatically permits a much reduced computational load. As an example, if the IIR filter is elliptic with order five, it can be implemented as a sum of two allpass functions [12] (of orders two and three respectively), requiring a total of only five multiplications per sample of $y(n)$! In addition, if we wish to find a decimated extrapolation, then these filters can be implemented in polyphase form [11] to improve efficiency.

Examples. Suppose we are given a 15 point sequence (i.e., $N = 15$; Fig. 1) which should be extrapolated to a 55 point sequence with $\sigma = \pi/3$. Suppose we wish to use an FIR filter for this. The length of the output is $M+15+K$ where K is the filter order. We shall assume the segment to be centered with respect to $y(n)$ (which, by construction, is causal) so that $M = K = (55-15)/2 = 20$. So we use a 20th order FIR filter and solve (13) recursively for increasing values of M (starting from $M = 0$), until we reach $M = 15$. Once we have designed the optimal (MNLS) input in this way, we run the FIR filter to compute the missing samples. Fig. 2(a) shows the energy of the input $u(n)$ as M increases; This is monotone decreasing as expected. Fig. 2(b) shows the percent of input energy in $|\omega| \geq \sigma$, which decreases as M increases beyond 5. The energy of the extrapolated output is plotted in Fig. 2(c) and again decreases as M increases (even though this cannot be formally proved in general). Fig. 2(d) shows the filter response $|H(e^{j\omega})|$ and finally Fig. 2(e) shows the plot of $|Y(e^{j\omega})|$ which verifies the 'bandlimited' nature of extrapolation $y(n)$. (These plots are normalized to 0 dB.)

For the IIR example, an elliptic lowpass filter with order five was first designed. This can be decomposed [12] into $(A_0(z) + A_1(z))/2$ where $A_0(z)$ and $A_1(z)$ are allpass with orders 3 and 2 respectively. To improve the condi-

tion number of the matrix A_M , a lifted filter [13] of the form $H(z) = (\cos \theta A_0(z) + \sin \theta A_1(z))/c$ was used with $\theta = 45.1$ degrees, and $c = \cos \theta + \sin \theta$. Fig. 3(a) shows the response $|H(e^{j\omega})|$ while Fig. 3(b) shows the quantity $|Y(e^{j\omega})|$ for the extrapolated output with $M = 20$. This plot is obtained with a windowed version of the infinitely long $y(n)$.

III. PHILOSOPHICAL DISCUSSIONS

From Sec. 1.1 we know that the discrete time σ -BL extrapolation problem has infinite number of solutions, so it is simply impossible to tell what the original bandlimited sequence was (i.e., where the segment $x(n)$ came from). Given an extrapolation $y_1(n)$ satisfying (1) we can find another extrapolation $y_2(n)$ satisfying (1) such that $y_1(n) - y_2(n)$ has arbitrarily large energy. And yet, $y_1(n)$ is as likely to be the original sequence as $y_2(n)$! In this sense, one can even argue that the extrapolation problem does not 'make sense'. One approach to overcome this difficulty is to impose the minimum norm condition so that a unique solution is obtained. (A second approach would be to take the solution to be the one for which the *normalized* distance from all other solutions is minimized [3].) But apart from the fact that it returns a unique solution, does this method have any further merit? This appears to require further thought. In any case, it does not seem appropriate to judge the merit of a method by checking how close the result is to an experimental, computer generated, 'original sequence' from which $x(n)$ has been obtained by truncation for testing purposes.

The finite-length extrapolation problem, on the other hand, gives a unique optimal solution because the quantity (12) is minimized by a unique extrapolation $y(n)$. The non uniqueness in this formulation is completely summarized by the single parameter α , which is a tradeoff between two quantities: (i) the degree to which $y(n)$ is bandlimited, and (ii) the fractional energy of $x(n)$ with respect to $y(n)$. Because of the optimal nature of the solution for fixed α , we can take the optimal solution to be a reference of comparison to the sub optimal methods described in Sec. 2.2.

To be more specific suppose we define

$$R_t = \sum_n |x(n)|^2 / \sum_n |y(n)|^2, \quad (18)$$

$$R_f = \int_{\sigma}^{\pi} |Y(e^{j\omega})|^2 d\omega / \int_0^{\pi} |Y(e^{j\omega})|^2 d\omega.$$

R_t and R_f are the time-domain and frequency-domain energy ratios, respectively. The aim, in general is to obtain large R_t (so that the original segment is not unduly dominated by the extrapolated part) and small R_f (so that $y(n)$ has small out-of-band energy). The tradeoff between these

two is effected by choice of α in the optimal method that minimizes (12). For the example presented in the previous section (FIR method which used an FIR filter of order 20), these ratios are $R_t = 0.39$ and $R_f = 0.000445$. For the optimal method, with α appropriately chosen we can obtain same R_f for $R_t = 0.69$. This also shows the degree to which the FIR method is suboptimal. Several other methods for finite length σ -BL extrapolation have been developed in [6], including windowed pseudoinverse methods, and fast recursive least squares (RLS) methods. The quantities R_t and R_f for all these methods are compared in [6], and it is found that the windowed pseudoinverse method is very close to optimal (even though slower than the FIR and IIR methods of Sec. 2.2).

References

- [1] A. K. Jain and S. Ranganath, "Extrapolation algorithms for discrete signals with application in spectral estimation," IEEE Trans. ASSP, pp. 830-845, Aug. 1981.
- [2] B. J. Sullivan and Bede Liu, "On the use of singular value decomposition and decimation in discrete-time band-limited signal extrapolation," IEEE Trans. ASSP, pp. 1201-1212, Dec. 1984.
- [3] D. P. Kolba and T. W. Parks, "Extrapolation and spectral estimation for bandlimited signals," Proc. of the ICASSP, pp. 372-374, April 1978.
- [4] A. Papoulis, *Signal Analysis*, McGraw Hill, 1977.
- [5] V. C. Liu and P. P. Vaidyanathan, "Finite length bandlimited extrapolation of discrete signals," Proc. of the ISCAS, pp. 1037-1040, May 1989.
- [6] P. P. Vaidyanathan and Vinay P. Sathé, "Fast techniques for band limited extrapolation of finite length sequences," Tech. Rep., Caltech, June 1989.
- [7] S. Haykin, *Adaptive filter theory*, Prentice Hall, 1986.
- [8] J. W. Adams and A. N. Willson, Jr., "A new approach to FIR digital filters with fewer multipliers and reduced sensitivity," IEEE Trans. CAS, pp. 277-283, May 1983.
- [9] Y. Neuvo, C.-Y. Dong, and S. K. Mitra, "Interpolated FIR filters," IEEE Trans. ASSP, pp. 563-570, June 1984.
- [10] M. H. Hayes and R. W. Schafer, "On the band-limited extrapolation of discrete signals," Proc. of the ICASSP, pp. 1450-1453, April 1983.
- [11] R. E. Crochiere and L. R. Rabiner, *Multirate digital signal processing*, Prentice Hall, Inc., 1983.
- [12] P. P. Vaidyanathan, et. al, "A new approach to the realization of low sensitivity IIR digital filters," IEEE Trans. ASSP, pp. 350-361, April 1986.
- [13] R. D. Koilpillai, et. al., "On arbitrary level IIR and FIR filters," IEEE Trans. on CAS, to appear.

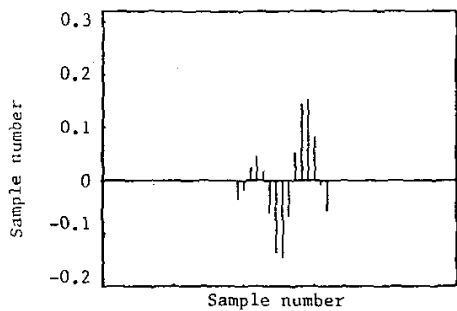


Figure 1. The segment to be extrapolated.

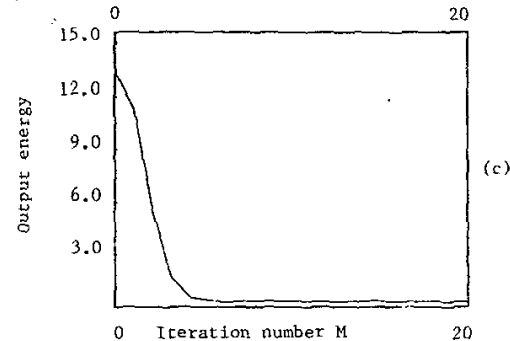
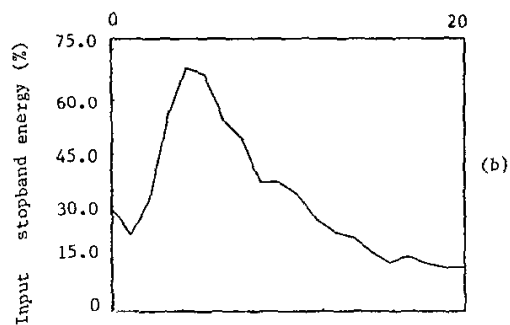
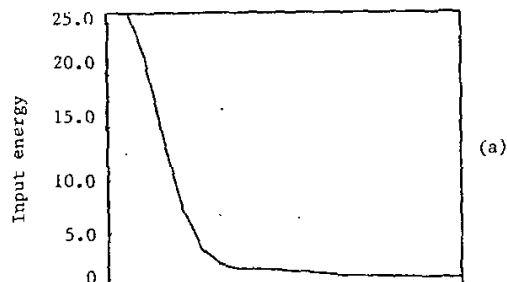


Figure 2. (a), (b), (c). Pertaining to FIR extrapolation example.

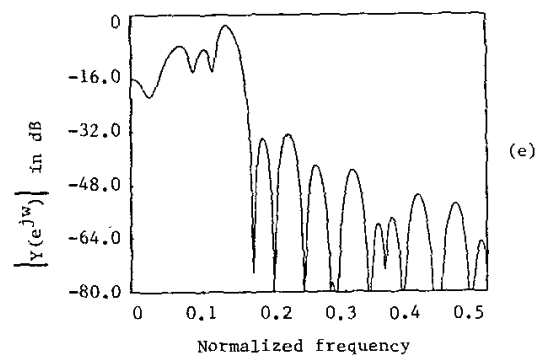
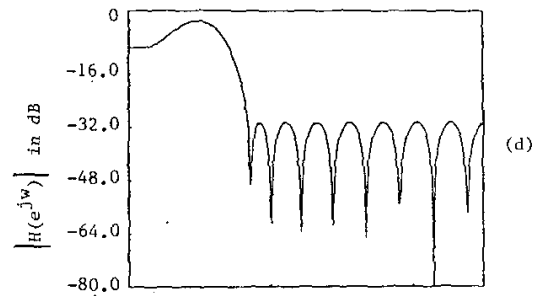


Figure 2. (d), (e), Pertaining to FIR extrapolation example.

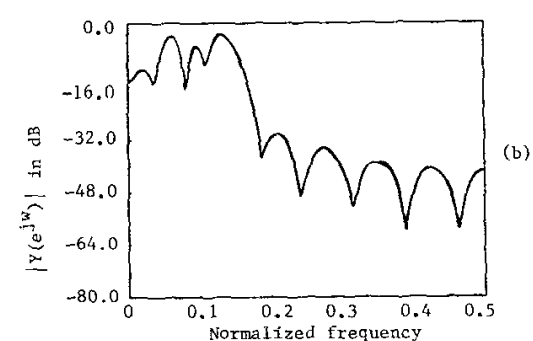
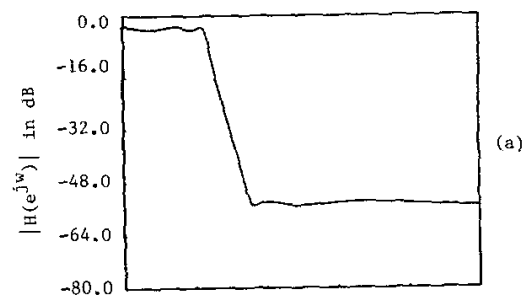


Figure 3. Pertaining to IIR extrapolation example.